Reference Resolution

Regina Barzilay
EECS Department
MIT
November 30, 2004

Reference Resolution: Example

The Salesgirl (Burns and Allen)

Gracie: And then Mr. and Mrs. Jones were having matrimonial trouble, and my brother was hired to watch Mrs. Jones.
George: Well, I am imagine she was a very attractive woman.
Gracie: She was, and my brother watched her day and night for six month.
George: Well, what happened?
Gracie: She finally got a divorce.
George: Mrs. Jones?
Gracie: No, my brother’s wife.

Task: determine which noun phrases refer to each real-world entity mentioned in a document

Goal: partition noun phrases in a text into coreference equivalence classes, with one cluster for each set of coreferent NPs

In the previous example:
{Mrs. Jones, she, she, Mrs. Jones},
{my brother, my brother},
{my brother’s wife}
Captain Farragut was a good seaman, worthy of the frigate he commanded. His vessel and he were one. He was the soul of it.

- Coreference resolution: {the frigate, his vessel, it}
- Anaphora resolution: {his vessel, it}

Coreference is a harder task!

Motivation

- Information extraction
- Question-Answering
- Machine-Translation

pronoun in the Malay language is translated by its antecedent (Mitkov, 1999)

- Summarization

When something goes wrong

In the past decade almost all Islamic revivalist movements have been labeled fundamentalists, whether they be of extremist or moderate origin. The widespread impact of the term is obvious from the following quotation from one of the most influential Encyclopedias under the title ‘Fundamentalist’: “The term fundamentalist has... been used to describe members of militant Islamic groups.” Why would the media use this specific word, so often with relation to Muslims? Most of them are radical Baptist, Lutheran and Presbyterian groups.
When something goes wrong

*Why would the media use this specific word, so often with relation to Muslims?*

*Before the term fundamentalist was branded for Muslims, it was, and still is, being used by certain Christian denominations. Most of them are radical Baptist, Lutheran and Presbyterian groups.*

**Types of referential expressions: Nouns**

- **Indefinite Noun Phrases:**
  - I saw an Acura Integra today.
  - Some Acura Integras were being unloaded.
  - I saw this awesome Acura Integra today.

- **Definite Noun Phrases**
  - I saw an Acura Integra today. The Integra was white and needed to be washed.
  - The fastest car in the Indianapolis 500 was an Integra.

**Pronouns**

Stronger constraints on using pronouns than on noun phrase references.

- Requires a high degree of activation from a referent
- Has a short activation span

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a. John went to Bob’s party, and parked next to a Acura Integra.</td>
<td>b. He went inside and talked to Bob for more than an hour.</td>
</tr>
<tr>
<td>a. Bob told him that he recently got engaged.</td>
<td>b. ??He also said that he bought it yesterday.</td>
</tr>
</tbody>
</table>

**Demonstratives and One Anaphora**

- Demonstratives (this, that) capture spatial proximity
  - I like this one, better than that

- One Anaphora evokes a new entity into the discourse whose description is dependent of this new entity
  - I saw no less that 6 Acuras today. Now I want one.
Troublemakers

- Inferables: inferential relation to an evoked entity
  I almost bought an Acura today, but a door had a dent and the engine seemed noisy.

- Discontinuous Sets: refer to entities that do not form a set in a text
  John has an Acura, and Mary has a Mazda. They drive them all the time.

- Generics: refer to general set of entities (in contrast to a specific set mentioned in text)
  I saw no less than six Acuras today. They are the coolest cars.

Syntactic Constraints

- Gender Agreement
  John has an Acura. It is attractive.

- Syntactic Agreement
  * John bought himself a new Acura.
  John bought him a new Acura.

Semantic Constraints

- Selectional restrictions of the verb on its arguments
  (1) John parked his Acura in the garage. He had driven it around for hours.
  (2) John parked his Acura in the garage. It is incredibly messy, with old bike and car parts lying around everywhere.
  (3) John parked his Acura in downtown Beverly Hills. It is incredibly messy, with old bike and car parts lying around everywhere.
Preferences in Pronoun Interpretation

- Recency: Entities introduced in recent utterances are more salient than those introduced further back
  John has an Integra. Bill has a Legend. Mary likes to drive it.

- Repeated mention: Entities that have been focused on in the prior discourse are more likely to continue to be focused on in subsequent discourse
  John needed a car to get his new job. He decided that he wanted something sporty. Bill went to the Acura dealership with him. He bought an Integra.

Preferences in Pronoun Interpretation

- Grammatical Role: Hierarchy of candidate entities based on their grammatical role
  John went to the Acura dealership with Bill. He bought an Integra.
  Bill went to the Acura dealership with John. He bought an Integra.

- Parallelism:
  Mary went with Sue to the Acura dealership. Sally went with her to the Mazda dealership.

Preferences in Pronoun Interpretation

Verb Semantics: emphasis on one of verb's arguments

- “implicit causality” of a verb causes change in salience of verb arguments
  John telephoned Bill. He lost the pamphlet on Acuras. John criticized Bill. He lost the pamphlet on Acuras.

- thematic roles (Goal, Source) cause change in salience of verb arguments
  John seized the Acura pamphlet from Bill. He loves reading about cars.
  John passed the Acura pamphlet to Bill. He loves reading about cars.

Hobbs’ Algorithm (1976)

- Features: Fully Syntactic
  - search the parse in a left-to-right, breadth-first fashion
  - give a preference to antecedent that are closer to the pronoun
  - give a preference to subjects

- Accuracy: 82%
Generic Algorithm

- Identification of Discourse Entities
  Identify nouns and pronouns in text

- Characterization of Discourse Entities
  Compute for each discourse entity NP<sub>i</sub> a set of values from \{K<sub>i1</sub>, . . . , K<sub>irm</sub>\} from \(m\) knowledge sources

- Anaphoricity Determination
  Eliminate non-anaphoric expressions to cut search space

- Generation of Candidate Antecedents
  Compute for each anaphoric NP<sub>j</sub> a list of candidate antecedents C<sub>j</sub>

Reference Resolution 20/55

Generic Algorithm (cont.)

- Filtering
  Remove all the members of C<sub>j</sub> that violate reference constraints

- Scoring/Ranking
  Order the candidates based on preferences and soft constraints

- Searching/Clustering
  Clustering of instances with the same antecedent

Reference Resolution 21/55

Reference Resolution: Trends

- Knowledge-Rich Approaches vs Knowledge-Lean Approaches

- Semi-automatic vs Fully-Automatic Preprocessing

- Small-scale vs Large-Scale Evaluation

Knowledge-Lean Multi-strategy Approach

(Lappin & Leass, 1994)

- Integrates the effects of the recency and syntactically-based preferences

- Doesn’t rely on semantic or pragmatic knowledge

- Follows greedy strategy

- Two stages: discourse model update and pronoun resolution

Reference Resolution 22/55

<table>
<thead>
<tr>
<th>Reference Resolution</th>
<th>21/55</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference Resolution</td>
<td>22/55</td>
</tr>
<tr>
<td>Reference Resolution</td>
<td>23/55</td>
</tr>
</tbody>
</table>
Discourse Model Update

(Lappin&Leass, 1994)

- Add every new discourse entity to discourse model
- Update its value based on salience factors
- Cut in half recency values when processing new entity (recency enforcement)

Syntactic Factors

subject > existential predicate nominal > object >
indirect object > demarcated adverbial PP

1. An Acura Integra is parked on the lot. (subject)
2. There is an Acura Integra parked in the lot.
3. ...
4. Inside his Acura Integra, John kissed Mary. (demarcated adverbial PP)

Penalty for non-head occurrences
Score for equivalence classes

Salience Factors

<table>
<thead>
<tr>
<th>Salience Factors</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentence Recency</td>
<td>100</td>
</tr>
<tr>
<td>Subject Emphasis</td>
<td>80</td>
</tr>
<tr>
<td>Existential Emphasis</td>
<td>70</td>
</tr>
<tr>
<td>Accusative</td>
<td>50</td>
</tr>
<tr>
<td>Indirect Object</td>
<td>40</td>
</tr>
<tr>
<td>Non-adverbial Emphasis</td>
<td>50</td>
</tr>
<tr>
<td>Head-noun Emphasis</td>
<td>80</td>
</tr>
</tbody>
</table>

Algorithm

1. Remove potential referents that do not agree in number or gender with the pronoun
2. Remove potential referents that do not pass intrasentetial syntactic coreference constraints
3. Update the total salience value of the referent
4. Select the referent with the highest value

Accuracy on unseen data: 86%
Clustering for Coreference
(Cardie&Wagstaff:1999)

- Each group of coreferent noun phrases defines an equivalence class
- Distance measure incorporates “linguistic intuition” about similarity of noun phrases
- Hard constraints enforce clustering construction

Distance Metric

\[
dist(NP_i, NP_j) = \sum_f w_f * incomp_f(NP_i, NP_j)
\]

Instance Representation

Based noun phrases (automatically computed) are represented with 11 features:

- Individual Words
- Head Word
- Position
- Pronoun type (nominative, accusative)
- Semantic Class: Time, City, Animal, Human, Object (WordNet)
- Gender (WordNet, specified list)
- Animacy (based on WordNet)

Clustering Algorithm

- Initialization: every noun is a singleton
- From right to left, compare each noun to all proceeding clusters
- Combine “close enough” clusters unless there exist any incompatible NP

Example: The chairman spoke with Ms. White. He...
Results

MUC-6 (30 documents): Recall 48.8%, Precision 57.4%, F-measure 52.8%
Baseline: 34.6%, 69.3%, 46.1%

Types of Mistakes:
- Parsing mistakes
- Coarse entity representation and mistakes in feature computation
- Greedy nature of the algorithm

Adding Linguistic Knowledge

Rich Linguistic representation for learning (Ng&Cardie 2002)
- 53 features
- manual feature selection
- significant gain in performance over (Soon et al., 2001)

Supervised Learning

(Soon et al., 2001)
- Decision Tree Induction
- Shallow feature representation (12 features):
  - “corrective” clustering
- Significant performance gain over rule-based algorithms

Features (Soon et al, 2001)
- distance in sentences between anaphora and antecedent?
- antecedent in a pronoun?
- weak string identity between anaphora and antecedent?
- anaphora is a definite noun phrase?
- anaphora is a demonstrative pronoun?
- number agreement between anaphora and antecedent
- semantic class agreement anaphora and antecedent
- gender agreement between anaphora and antecedent
- anaphora and antecedent are both proper names?
- an alias feature
- an appositive feature
Observations

(Ng&Cardie’2002)

- Feature selection plays an important role in classification accuracy: MUC-6 62.6% (Soon et al., 2001) → Ng&Cardie, 2002) 69.1%
- Clustering operates over the results of hard clustering, which may negatively influence the final results
- Machine learning techniques rely on large amounts of annotated data: 30 texts
- All the methods are developed on the same corpus of newspaper articles

Classification Rules

```
+ 786 59 IF SOON-WORDS-STR = C
+ 73 10 IF WNCLASS = C PROPER-NOUN = D NUMBERS = C SENTNUM <= 1 PRO-RESOLVE = C ANIMACY = C
+ 40 8 IF WNCLASS = C CONSTRAINTS = D PARANUM <= 0 PRO-RESOLVE = C
+ 16 0 IF WNCLASS = C CONSTRAINTS = D SENTNUM <= 1 BOTH-IN-QUOTES = 1 APPOSITIVE = C
+ 17 0 IF WNCLASS = C PROPER-NOUN = D NUMBERS = C PARANUM <= 1 BPRONOUN-1 = Y AGREEMENT = C CONSTRAINTS = C BOTH-PRONOUNS = C
+ 38 24 IF WNCLASS = C PROPER-NOUN = D NUMBERS = C SENTNUM <= 2 BOTH-PRONOUNS = D AGREEMENT = C SUBJECT-2 = Y
+ 36 8 IF WNCLASS = C PROPER-NOUN = D NUMBERS = C BOTH-PRONOUNS = C
+ 11 0 IF WNCLASS = C CONSTRAINTS = D SENTNUM <= 3 SUBJECT-1 = Y SUBJECT-2 = Y SUBCLASS = D IN-QUOTE-2 = N BOTH-DEFINITES = 1
```

Co-training

(Blum&Mitchell, 1998)

1. Given a small amount of training data, train two classifiers based on orthogonal set of features
2. Add to training set $n$ instances on which both classifiers agree
3. Retrain both classifiers on the extended set
4. Return to step 2
Co-training for Coreference

Coreference does not support natural split of features
Algorithm for feature splitting

- Train a classifier on each feature separately
- Select the best feature and assign it to the first view, and the second best feature assign to the second view
- Iterate over the remaining feature, and add them to one of the views

Separate training for each reference type (personal pronouns, possessives, . . .)

Reference Resolution 40/55

Results

Improvements for some types of references

- Definite noun phrases: from 19% to 28% (2000 training instances)
- No improvements for possessives, proper names and possessive pronouns

Study of learning curves

- Personal and possessive pronoun can be trained from very small training data (100 instances)
- Other types of references require large amounts of training data

Reference Resolution 42/55

Anaphora In Spoken Dialogue

Differences between spoken and written text

- High frequency of anaphora
- Presence of “Vague anaphora” (Eckert&Strube'2000) 33%
- Presence of non-NP-antecedents (Byron&Allen’1998) TRAINS93: 50% (Eckert&Strube’2000) SwitchBoard: 22%
- Presence of repairs, disfluences, abandoned utterances and so on...

Reference Resolution 43/55

Example of Dialog

A1: ..[he]’s nine months old . . .
A2: ..[He], likes to dig around a little bit.
A3: ..[His mother]’s mother comes in and says, why did you let [him] [plays in the dirt].
A4: I guess [[he]’s enjoying himself].
B5: [That]’s right.
B6: [It]’s healthy . . .
(Webber, 1988)
(A0) Each Fall, penguins migrate to Fiji.
(A1) That’s where they wait out the winter.
(A2) That’s when it’s cold even for them.
(A3) That’s why I’m going there next month.
(A4) It happens just before the eggs hutch.

Reference Resolution

Symbolic Approach

Pronominal Anaphora Resolution (Byron, 2002)
- Mentioned Entities — referents nouns phrases
- Activated Entities — entire sentences and nominals
- Discourse Entity attributes:
  - Input: The surface linguistic constituent
  - Type: ENGINE, PERSON, ...
  - Composition: hetero- or homogeneous
  - Specificity: individual or kind

Abstract Referents

- Webber (1990): each discourse unit produces a pseudo discourse entity — “proxy for its propositional content”
- Abstract Pronoun interpretation: requires presentation of fact referents
- Walker&Whittaker (1990): in problem-solving dialogs, people refer to aspects of the solution that were not explicitly mentioned (Byron, 2002)
  A1 Send engine to Elmira.
  A2 That’s six hours.

Activated Entities

Generation of Multiple Proxies

- To load the boxcars/Loading them takes an hour (infinitive or gerund phrase)
- I think he that he’s an alien (the entire clause)
- I think that he’s an alien (sentential)
- If he’s an alien (Subordinate clause)
Types of Speech Acts

Tell, Request, Wh-Questions, YN-Question, Confirm
(1) The highway is closed (Tell)
(2) Is the highway closed? (Y/N Question)
(3) That's right.
(4) Why is the highway closed? (WH-Q)
(5) *That's right.

Reference Resolution 48/55

Semantic Constraints

“Heavily-typed” system
- Verb Senses (selectional restrictions)
  “Load them into the boxcar” (them has to be CARGO)
- Predicate NPs
  “That’s a good route“ (that has to be a ROUTE)
- Predicate Adjectives
  “It’s right” (it has to be a proposition)

Evaluation

10 dialogues, 557 utterances, 180 test pronouns
- Salience-based resolution: 37%
- Adding Semantic constraints: 43%
- Adding Abstract referents: 67%
- “Smart” Search order: 72%
- Domain Independent Semantics: 51%

Example

Engine 1 goes to Avon to get the oranges.
(TELL (MOVE :theme x :dest y :reason (LOAD :theme w)))
(the x (refers-to x ENG1))
(the y (refers-to y AVON))
(the w (refers-to w ORANGES))
So it'll get there at 3 p.m.
(ARRIVE :theme x :dest: y :time z)
“get there” requires MOVABLE-OBJECT

Reference Resolution 50/55

Reference Resolution 51/55
Knowledge-Lean Approach

(Strube&Müller’2003)

- Switchboard: 3275 sentences, 1771 turns, 16601 markables
- Data annotated with disfluency information
- “Problematic” utterances were discarded
- Approach: ML combines standard features with dialogue specific features

Features

Features induced for spoken dialogue: ante-exp-type [type of antecedent (NP, S, VP)]
ana-np-pref [preference for NP arguments]
mdist-3mf3p [the number of NP-markables between anaphora and potential antecedent]
ante-tfidf [the relative importance of the expression in the dialogues]
average-ic [information content: neg. log of the total frequency of the word divided by number of words]

F-measure:
- Fem&Masc Pronoun: 17.4% baseline, 17.25%
- Third Person Neuter Pronoun: 14.68%, 19.26%
- Third Person Plural: 28.30%, 28.70%

Observations

- Coreference for speech processing is hard!
- New features for dialogue are required
- Prosodic features seems to be useful